CLAIMS

WHAT IS CLAIMED IS:

1.   A method of inferring a function of one or more genes, the method comprising:

assembling a distribution of gene expression level measurements, wherein each of said gene expression level measurements comprises at least one parameter;

calculating a significance score for each gene expression level measurement in said distribution;

selecting at least one gene expression level measurement from said distribution based on its significance score;

comparing said at least one parameter of said selected at least one gene expression level measurement to said at least one parameter of non-selected gene expression level measurements; and

determining said at least one parameter from said selected at least one gene expression level measurement that is not present in said non-selected gene expression level measurements.

2.   The method of claim 1, wherein said gene expression level measurements are accessed from a database comprising gene expression level measurements for a plurality of genes.

3.   The method of claim 1, wherein said significance score indicates a measure of the distance said gene expression level measurements are from a mean value of said gene expression level measurements.

4.   The method of claim 3, wherein the significance score is a z-score.

5.   The method of claim 4, wherein selected said at least one gene expression level measurement has a z-score greater than 3 and less than –3.

6.   The method of claim 3, wherein said at least one gene expression level measurement is selected based on its significance score being a predefined distance from said mean value of said gene expression level measurements.

7.   The method of claim 1, wherein said at least one parameter comprises at least one of ecotype, tissue, RNA type, harvest conditions, genotype, growth conditions, growth media, treatments, and combinations therein.

8.   The method of claim 1, further comprising sorting the expression level measurements based on said significance scores of said gene expression level measurements.

9.   The method according to claim 1, further comprising inferring the function of said one or more genes based on said at least one parameter from said selected at least one gene expression level measurement determined not to be present in said non-selected gene expression level measurements.

10.   A computer readable medium having instructions for causing a computer to perform a method of inferring a function of one or more genes, the method comprising:

        assembling a distribution of gene expression level measurements, wherein each of said gene expression level measurements comprises at least one parameter;

calculating a significance score for each gene expression level measurement in said distribution;

selecting at least one gene expression level measurement from said distribution based on its significance score;

comparing said at least one parameter of said selected at least one gene expression level measurement to said at least one parameter of non-selected gene expression level measurements; and

determining said at least one parameter from said selected at least one gene expression level measurement that is not present in said non-selected gene expression level measurements.

11. The computer readable medium claim 10, wherein said gene expression level measurements are accessed from a database comprising gene expression level measurements for a plurality of genes.

12. The computer readable medium of claim 10, wherein said significance score indicates a measure of the distance said gene expression level measurements are from a mean value of said gene expression level measurements.

13. The computer readable medium of claim 12, wherein the significance score is a z-score.

14. The computer readable medium of claim 13, wherein selected said at least one gene expression level measurement has a z-score greater than 3 and less than − 3.

15.   The computer readable medium of claim 12, wherein said at least one gene

expression level measurement is selected based on its significance score being a

predefined distance from said mean value of said gene expression level

measurements.

16.   . The computer readable medium of claim 10, wherein said at least one

parameter comprises at least one of ecotype, tissue, RNA type, harvest conditions,

genotype, growth conditions, growth media, treatments, and combinations therein.

17.   The computer readable medium of claim 10, further comprising sorting the

expression level measurements based on said significance score of said gene

expression level measurements.

18.   A system for inferring a function of one or more genes, the method

comprising:

        means for assembling a distribution of gene expression level measurements,

wherein each of said gene expression level measurements comprises at least one

parameter;

        means for calculating a significance score for each gene expression level

measurement in said distribution;

        means for selecting at least one gene expression level measurement from

said distribution based on its significance score;

        means for comparing said at least one parameter of said selected at least one

gene expression level measurement to said at least one parameter of non-selected

gene expression level measurements; and

means for determining said at least one parameter from said selected at least one gene expression level measurement that is not present in said non-selected gene expression level measurements.

19. The system of claim 18, wherein said gene expression level measurements are accessed from a database comprising gene expression level measurements for a plurality of genes.

20. The system of claim 18, wherein said significance score indicates a measure of the distance said gene expression level measurements are from a mean value of said gene expression level measurements.

21. The system of claim 20, wherein said significance score is a z-score.

22. The system of claim 21, wherein said at least one gene expression level measurement selected has a z-score greater than 3 and less than –3.

23. The system of claim 20, wherein selected said at least one gene expression level measurement is based on its significance score being a predefined distance from said mean value of said gene expression level measurements.

24. The system of claim 18, wherein said at least one parameter comprises at least one of ecotype, tissue, RNA type, harvest conditions, genotype, growth conditions, growth media, treatments, and combinations therein.

25. The system of claim 18, further comprising sorting the expression level measurements based on said significance scores of said gene expression level measurements.

26. The system according to claim 18, further comprising inferring the function of said one or more genes based on said at least one parameter from said selected at least one gene expression level measurement determined not to be present in said non-selected gene expression level measurements.

27. A computer system useful for inferring functions of genes, the computer system comprising:

a processor;

a memory coupled to the processor;

a display coupled to the processor;

a computer program that executes on the processor, the program further comprising:

a module assembling a distribution of gene expression level measurements, wherein each of said gene expression level measurements comprises at least one parameter,

a module calculating a significance score for each gene expression level measurement in said distribution,

a module selecting at least one gene expression level measurement from said distribution based on its significance score,

a module comparing said at least one parameter of said selected at least one gene expression level measurement to said at least one parameter of non-selected gene expression level measurements, and

a module determining said at least one parameter from said selected at least one gene expression level measurement that is not present in said non-selected gene expression level measurements.

28. A method of inferring a function of one or more genes, the method comprising:

accessing multiple expression level measurements for multiple samples collected under diversified conditions defined by parameters corresponding to a gene;

calculating a score for the expression level measurements;

selecting samples based on most significant score;

determining non-overlapped parameters between the selected samples and the sample that were not selected; and

inferring gene functions from the non-overlapped parameters.

29. A method for inferring the function of a biological molecule, said method comprising:

assembling a distribution of data points from experimental measurements, wherein each of said data points comprises at least one parameter;

calculating a significance score for each data point in said distribution;

selecting at least one of said date points based on its significance score;

comparing said at least one parameter of said selected at least one data point to said at least one parameter of non-selected data points; and

determining said at least one parameter from said selected at least one data point that is not present in said non-selected data points.

30. The method according to claim 29, further comprising inferring the function of said biological molecule based on said at least one parameter from said selected at least one data point determined not to be present in said non-selected data points.